Choosing an Interim Assessment: Guidelines for Stakeholders

November 2018

Choosing an Appropriate Interim Assessment:
Guidelines for Stakeholders

WIDA stakeholders have increasingly expressed interest in interim assessments that are linked to ACCESS for ELLs.[1] The value that interim assessments contribute to a comprehensive assessment system is well established (Herman, 2017; Perie, Marion, Gong, & Wurtzel, 2007); they can provide useful information to inform instruction, gauge student progress, and predict summative assessment performance. It is worth noting that interim assessments are typically designed with information on "unit attainment or progress across units. Interim assessment can be seen as short-cycled summative assessments" (MacDonald, Boals, Castro, Cook, Lundberg, and White, 2015, p.xvi).

However, before assessments can be identified and used as interim assessments for any context, including the context in which ACCESS for ELLs is administered, evidence should be provided that shows that they are appropriate for the purpose. This document describes WIDA's viewpoint of the criteria for an assessment to be used as a valid interim English language proficiency assessment in support of the annual administration of ACCESS for ELLs.

Approach: Two resources were used to assist in outlining critical categories: the literature about important criteria for validating interim assessments and contextual framing of how this type of interim assessment would be used.

Critical Categories
In order for an assessment to be used as an interim assessment for ACCESS for ELLs, a validity argument must be made. The critical categories below were selected on the basis of their role in demonstrating the valid use of an interim assessment for ACCESS for ELLs test-takers.

1. Theoretical foundation
A foundation in second language acquisition/development theory that is compatible with the WIDA Guiding Principles of Language Development and Learning for Multilingual Learners (https://wida.wisc.edu/sites/default/files/resource/Guiding-Principles-of-Language-Development.pdf) and the WIDA Standards Framework (https://wida.wisc.edu/sites/default/files/resource/WIDA-Standards-Framework-and-its-Theoretical-Foundations.pdf) is necessary. For an interim assessment to be suitable for use in conjunction with ACCESS for ELLs, it must have been designed for the same use, and be able to demonstrate what approach to English language development undergirds the assessment.

Example of evidence:
- Formal test documents describing the theoretical framework within which the test was designed and content and tasks were developed.

---

[1] ACCESS for ELLs is a secure, large-scale English language proficiency assessment that is administered to Kindergarten through 12th grade students who have been identified as English language learners. The assessment is given annually in WIDA consortium member states and territories in order to monitor students' progress in acquiring academic English language and students' readiness to participate meaningfully in content area classrooms without language instruction educational program support.

2. Content validity

ACCESS for ELLs is designed to test the breadth of the six proficiency levels of the five WIDA ELD standards (https://wida.wisc.edu/teach/standards/eld) with a balance of content by standard across the four domains of speaking, listening, reading, and writing. For an interim assessment to be used in conjunction with ACCESS for ELLs, it must show that it addresses a similar range of proficiency / developing language ability, and must provide a similar balance of content across the five WIDA standards across the four domains. Additionally, items and tasks that are developed for the assessment should all go through review steps that help ensure the content and cognitive match between material on the test and the proficiency level and age of test-takers.

a. Linguistic complexity
A range of proficiency within each grade and within each domain that is similar to ACCESS for ELLs.

b. Standards and balance of content
A balance of standards coverage across all grades and domains**.**

Example of evidence:
- A formal, independent alignment study that shows linguistic difficulty level and standards alignment.
- A test development report that provides a detailed description of the test content's development and review steps.

3. Technical quality

The technical quality of an interim assessment must meet industry standards for reliability and measurement error in order for it to be considered for use with ACCESS for ELLs. Demonstration of evidence of an internal structure that is similar to ACCESS for ELLs could also provide support for the technical quality of an interim assessment (and for content validity in critical category 2 above). Technical quality extends to the scoring of productive skills of speaking and writing -- documentation on the processes and procedures for ensuring high-quality rating must be provided (i.e., descriptions of training materials, rater qualifying, rater monitoring).

Example of evidence:
- Technical report providing the assessment's reliability estimates by domain and grade, including errors of measurement and confidence intervals.
- Reports of the underlying structure of tested construct (e.g., IRT fit or factor analysis) and correlations to similar types of constructs.
- Reports providing details on scoring processes and procedures. For constructed reponse assessments and/or items, reports and guidelines on test rubrics and rater training and scoring materials.
- Reports on how the interim assessment characterizes growth and how growth is unique across grade and language proficiency levels.

4. Score reports and interpretations

Score reports that provide clear and accessible information in a timely manner and supporting documentation for interpreting and using the score reports are essential for an effective interim assessment. Individual student score reports, easily understood by parents, must be provided, with

additional score reports that provide aggregate information at relevant levels (e.g., by a class, school or district level).

Example of evidence:
- A variety of score reports that are available to users and the interpretive guides or support materials that are available to help test users in interpreting scores.
- A score report on growth charcteristics of students at different grades and language proficiency levels.

5. Practicality and Inclusivity
The category of inclusivity is broken into three sub-categories: 1) the practicality of the assessment in general (how feasible is administration of the assessment, whether it is paper-based or computer-based, for the target student population), 2) fairness, and 3) accessibility and available accommodations for EL's with disabilities.

a. Practicality
Whether the test is administered on paper or on a computer, information should be provided on how students are registered, how schools/classes receive access to test materials, what kind of physical environment is necessary for administering the test, how long the test takes, etc. All operational aspects of administering the test must be provided.

Example of evidence:
- Test administration manuals that provide details to test administrators on how to administer the test.

b. Fairness
All assessments must be created with awareness that upsetting, disturbing or biased test content may affect test-taker performance in construct-irrelevant ways.

Example of evidence:
- Test development report including explanation of test development processes that are carried out to ensure minimization of potential fairness issues in test content.

c. Accessibility and accommodations
Test development best practice entails the design of assessments with universal design principles in mind. That is, to provide support to all test-takers in ways that allow them to demonstrate their full abilities. These kinds of support extend from the way test content is formatted to enhance the testing experience for all test-takers to the provision of alternative formats that are available to test takers with physical or cognitive differences (see http://cats.cuny.edu/events/CAC2010/Universal%20Design/Checklist%20for%20Universal%20Design%20of%20Tests.pdf).

Example of evidence:
- Explanation of accessibility tools and accommodations that are available to students.

6. Relationship to performance on ACCESS for ELLs

A key use of interim assessments is their predictive relationship with summative assessments that are part of an overall assessment program or system. It is therefore essential that interim assessments considered for use with ACCESS for ELLs provide evidence of a relationship with ACCESS in the key areas of growth and prediction across language domains and grades/grade clusters.

Example of evidence:
- Reports of statistical analyses (e.g., regression, hierarchical linear modeling) indicating predictive ability of the interim assessment to ACCESS scores or showing similar growth patterns.
- Reports showing the relationship between ACCESS proficiency levels and the interim assessment's proficiency levels.

7. Consequential validity: positive influence on classroom practices

A key use of an interim assessment is their ability to inform curriculum decisions and classroom practice, with great importance placed on the match between the language and types of tasks assessed and used on the assessment and the language and activities that exist within classrooms.

Example of evidence:
- Testimonials of how the interim assessment was used positively in classrooms, schools or districts contexts.
- Reports or studies of focus groups and/or surveys of teachers/administrators highlighting the positive influences on EL's English language proficiency development based on the assessment.

Conclusion

Interim assessments can play an important role in supporting learning and growth of students, but their fitness for use must be carefully considered. For an interim assessment to be useful in the context of WIDA ACCESS for ELL, the seven categories described above provide guidance on what should be considered essential, and WIDA believes that interim assessments that provide sufficient evidence in these categories will likely be able to provide curriculum and evaluation support that is beneficial to students in their ultimate goal of acquiring proficient academic English language skills.

Test publishers are encouraged to provide evidence in these areas to test users, and test users are encouraged to ask for this type of information. Table A1 in the Appendix is a template that test publishers can use in providing interim assessment evidence.

Question Checklist

Table 1 shows a sample of questions test users may want to ask test publishers of their prospective interim assessment, specifically as it relates to WIDA standards and ACCESS for ELLs:

Table 1: Guideline's Questions on Interim Assessments

Questions	Met Category Requirement		
	Yes	Partially	No
Critical category 1: Has the test publisher provided an explanation of the theoretical approach to language development that the test is grounded in?			
Critical category 2: Has the test publisher shown how the test is aligned to the WIDA ELP standards across grade clusters both in terms of content and proficiency levels?			
Critical category 3: Has the test publisher provided evidence of the statistical reliability of their test?			
Critical category 4: Are score reports available for the test and is there sufficent interpretative materials for score reports?			
Critical category 5: Has the test publisher provided a test administration manual that provides detailed information pertaining to the logistics of administering the test?			
Critical category 6: Has the test publisher conducted analyses to demonstrate the relationship of the test to ACCESS in terms of prediction and growth patterns?			
Critical category 7: Is there evidence provided to show the positive influence of the test on classroom practices?			

References

Board of Regents of the University of Wisconsin System (2010). The Cornerstone of the WIDA Standards: Guiding Principles of Language Development. Madison, WI: University of Wisconsin System. Retrieved from: https://wida.wisc.edu/sites/default/files/resource/Guiding-Principles-of-Language-Development.pdf

Board of Regents of the University of Wisconsin System (2012). *2012 Amplification of The English Language Development Standards*. Madison, WI: University of Wisconsin System.

Herman, J. (2017). Interim assessments in brief. California, USA: The Regents of the University of California. Retrieved from http://www.csai-online.org/sites/default/files/resources/4666/InterimAssessmentsBrief.pdf

MacDonald, R., Boals, T., Castro, M., Cook, H.G., Lundberg, T., and White, P.A. (2015). *Formative Language Assessment for English Learners: A four-step process*. Portsmouth, NH: Heinemann.

Perie, M., Marion, S, Gong, B., & Wurtzel, J. (2007). The role of interim assessments in a comprehensive assessment system: A policy brief. Washington, DC: Aspen Institute. Retrieved from https://www.achieve.org/files/TheRoleofInterimAssessments.pdf

Rehabilitation Design & Disability Center (2009). Checklist for Universal Design of tests. Milwaukee, WI: University of Wisconsin-Milwaukee.

Appendix

Table A1: Vendor Response to Interim Assessment Guidelines Associated with WIDA Standards and ACCESS for ELLs

Critical Category	Example Evidence	Supporting document(s)
1. Theoretical foundation	• Formal test documents describing the theoretical framework within which the test was designed and content and tasks were developed.	
2. Content validity	• A formal alignment study that shows linguistic difficulty level and standards alignment • A test development report that provides a detailed description of the test content's development and review steps	
3 .Technical quality	• Technical report providing the assessment's reliability estimates by domain and grade, including errors of measurement and confidence intervals • Reports of the underlying structure of tested construct (e.g., IRT fit or factor analysis) and correlations to similar types of constructs • Reports providing details on scoring processes and procedures. For constructed reponse assessments and/or items, reports and guidelines on test rubrics and rater training and scoring materials. • Reports on how the interim assessment characterizes growth and how growth is unique across grade and language proficiency levels	
4. Score reports and score interpretations	• A variety of score reports that are available to users and the interpretive guides or support materials that are available to help test users in interpreting scores. • A score report on growth charcteristics of students at different grades and language proficiency levels	
5. Practicality and Inclusivity	• Test administration manuals that provide details to test administrators on how to administer the test	

Table A1: Vendor Response to Interim Assessment Guidelines Associated with WIDA Standards and ACCESS for ELLs

Critical Category	Example Evidence	Supporting document(s)
	• Explanation of accessibility tools and accommodations that are available to students	
6. Relationship to performance on ACCESS for ELLs	• Reports of statistical analyses (e.g. regression) indicating predictive ability of the interim assessment to ACCESS scores or showing similar growth patterns • Reports showing the relationship between ACCESS proficiency levels and the interim assessment's proficiency levels	
7. Consequential validity	• Testimonials of how the interim assessment was used positively in classrooms, schools or districts contexts • Reports or studies of focus groups and/or surveys of teachers/administrators highlighting the positive influences on EL's English language proficiency development based on the assessment	